DOCUMENT RESUME

ED 218 356 TM 820 405

AUTHOR Klibanoff, Leonard S.; Gamel, Nona N.
TITLE Reallocating Students after the Pretest: Biasing the Norm-Referenced Model.
PUB DATE Mar 82
NOTE 36p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Gains; Compensatory Education; *Educational Testing; *Grouping (Instructional Purposes); Mathematical Models; Pretests Posttests; Primary Education; *Program Evaluation; Teacher Attitudes; *Test Reliability
IDENTIFIERS *Evaluation Problems; *RMC Models; Title I Evaluation and Reporting System

ABSTRACT

A large-scale and a small-scale simulation of the reallocation of students into and out of an originally-selected treatment group were performed. It was postulated that both the scores on a selection test and teacher ratings are measures of student needs in the Title I Evaluation and Reporting System (TIERS). In the norm-referenced model, three conditions of reallocation with three separate gains estimates were studied using second grade students data from the Study of the Sustaining Effects of Compensatory Education. Gain estimates obtained under the various reallocation conditions were compared to the gain that would have been obtained under correct implementation of this variation of the model of selection on the pretest with a strict cutoff and correction for regression. Gain scores in the large-scale simulation after reallocation differed from the benchmark by as much as .2 standard deviation units. The small-scale analysis showed that large biases can result at the local level if even one student is moved. (Author/CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED218356

REALLOCATING STUDENTS AFTER THE PRETEST: BIASING THE NORM-REFERENCED MODEL

Leonard S. Klibanoff
Nona N. Gamel

RMC Research Corporation
Mountain View, CA 94040

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

L.S. Klibanoff

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at
The Annual Meeting of the
American Educational Research Association
March 19-23, 1982
New York, NY

TM 820 405

Reallocating Students after the Pretest: Biasing the Norm-Referenced Model

The norm-referenced model has been the most widely used evaluation model in the Title I evaluation and reporting system (TIERS). The generation of the no-treatment expectation in this model is based on the equipercentile assumption, i.e., that without any special educational intervention, students will tend to maintain their position with respect to their peers throughout the school year. Until recently, one requirement of this model was that Title I students not be selected on the basis of their pretest scores. The use of a selection measure separate from the pretest was intended to avoid biased gain estimates which would otherwise result from regression to the mean.

Since it appeared that the usefulness of the norm-referenced model would be enhanced if it were possible to assign students to Title I projects on the basis of their pretest scores, Roberts (1980) gives the formula to correct for the bias introduced by selecting on the pretest. His analyses suggested that the following general formula provides an adequate correction when assignment to the Title I group is determined solely on the basis of pretest scores:

$$\bar{X}_{cs} = r_{xx} (\bar{X}_s - \bar{X}_g) + \bar{X}_g \qquad \text{----------(1)}$$

where

$\bar{X}_s$ = mean pretest score of the project students,

$\bar{X}_g$ = mean pretest score of the total group from which project students are selected,

r_{xx} = test-retest reliability for the total group from which project students are selected,

$\bar{X}_{cs}$ = corrected mean pretest score of the project students.

The introduction of this correction formula into the evaluation and reporting system offered users a choice between using the pretest score alone to select Title I participants (strict cutoff), and selecting students on the basis of some other measure or combination of measures. There was concern, however, that users would choose the option of selecting on the pretest score but then reassign certain students based on teacher judgments or other procedures. This phenomenon has been labelled the fuzzy cutoff. This concern is based on the knowledge that Title I personnel often are reluctant to base student selection solely on achievement test scores, since such scores are felt to be imperfect indices of true achievement levels. For example, data from the Sustaining Effects Study (Wang, Hoepfner, Zagorski, Hemenway, Brown, & Bear, 1978) indicate that about 44% of the districts in the study used teacher judgment equally weighted with test scores to select Title I participants (p. 207).[1] If, because of teacher judgment, certain lower scoring students are not selected for Title I services and other higher scoring students are selected, the correction formula presented above may no longer be adequate. Where the reallocation is valid (students with artificially high scores are moved into the group and students with artificially low scores moved out) the correction equation would be expected to lead to underestimates of project impact. It is also possible that the allocations of selected and nonselected students will result in overestimates of achievement gain. Such could be the case if the reallocation (and the ratings upon which they are based) were on some basis other than actual need for Title I services (e.g., behavioral problems).

---

[1]Note that the fuzzy cutoff does not necessarily correspond to the case of selection on a composite measure. If a composite (the elements entering the composite must be independent of the pretest) is used and a sharp cutoff (on the composite) is maintained, a separate pretest is required to implement the norm-referenced model.

In this study, the effects of reallocating students in conjunction with the correction formula is investigated as a function of the nature of the teacher judgment used to reassign students. Teacher judgment is modeled by postulating that both scores on a selection test and ratings by teachers (or other administrative procedures) are measures of students' needs for special educational service such as Title I. The two measures may be in more or less agreement with each other. The criterion for validity used in this study is purely a predictive one, the student's eventual performance on the posttest score. Measurement error on the posttest score are presumably uncorrelated with those on the pretest score. Hence, the valid case under this model may be seen as one in which the judgment has a low correlation with the pretest and a high correlation with the posttest. The invalid case is one in which the judgment has a high correlation with the pretest and a low correlation with the posttest, and the neutral case is one in which the two correlations are nearly equal. Teacher judgment is simulated using random numbers in conjunction with test scores (see below).

Method

Data Base

Analyses were conducted using data from the Study of the Sustaining Effects of Compensatory Education (Sustaining Effects Study or SES). Students in grades 2 through 6 in the SES were administered reading and mathematics portions of two adjacent levels of the Comprehensive Test of Basic Skills (CTBS) in both the fall and spring of the first year of the study. The data analysis file used in this study comprised a nationwide sample of approximately 6,800 students per grade who were receiving no compensatory education services. Each student had a total reading and a total math score for two test levels of the CTBS (at-level and below-level) for both fall and spring. For each grade within each school, System Development Corporation, which conducted the SES, determined the

"best" level of the tests, i.e., the level least likely to produce floor or ceiling effects.

Procedure

Subjects. The amount of bias introduced by both selection on the pretest (prior to employing the correction for regression-effect error (Roberts, 1980)) and by the reallocation of students after the pretest (i.e., the fuzzy cutoff) is a function of the reliability of the measures. Hence, our analyses were conducted at grade 2, wherein both the test scores and the judgments (one would suppose) exhibit the lowest levels of reliability of any grade level in the data base. In effect, grade 2 represents the "worst case" condition, enabling us to identify reasonable boundaries on the size of the bias. A group of students who were not participating in compensatory education programs and who had scores below 36.5 normal curve equivalents (NCEs) on the fall best-level reading test was selected from the 6,449 students on the file. This group includes roughly the bottom third of the score distribution and thus can be said to resemble a Title I group. This group became the simulated treatment group (referred to as the selected group) and all students with scores above 36.5 NCEs were in the nonselected group.

Simulated teacher judgment. For each student in both the selected and nonselected groups, a simulated teacher judgment variable was created using a linear composite of pretest score, posttest score, and a normally distributed random component. The random component was used to create specific correlations between the teacher judgment variable and the pretest and posttest scores for the group. A matrix of correlations was created to cover the range of neutral, valid, and invalid reallocations of students to the selected and nonselected groups based on the simulated teacher judgment (see Figure 1).

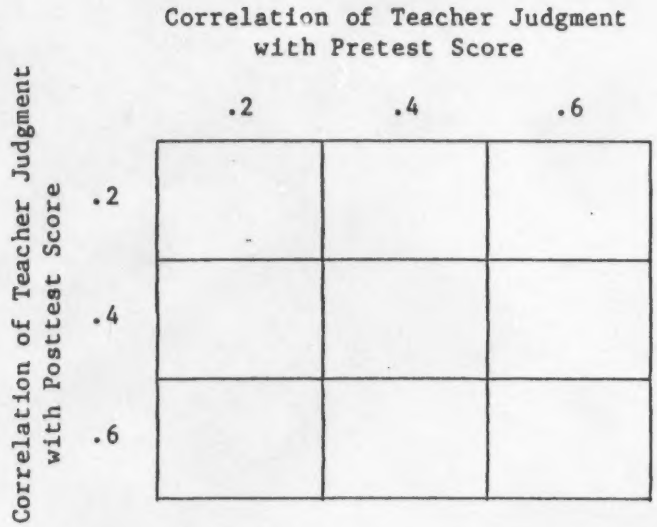

Figure 1. Target correlation of the teacher judgment (reallocation) variable with pretest and posttest scores.

For selected cells in the matrix (i.e., .2, .6; .6, .6; and .6, .2), the five percent of selected students with the highest simulated teacher judgment score were removed and placed in the nonselected group. These students were replaced with an equal number of nonselected students with the lowest teacher judgment score. This procedure was then repeated for each cell reallocating ten percent rather than five percent of the students from the selected group and an equal number of students from the nonselected group.

It should be clear that the methodology employed in this study is at least a bit unusual in that both empirical and simulated data are used jointly in the analyses. This approach is employed because we believed it to be important for the results to reflect both the range of effects caused by reallocating students where the same test is used for selection and pretest purposes and be veridical with respect to post-on-pretest regression functions in the real world. This choice of method

is not without some cost, however. By using empirical pretest and posttest scores with their existing correlation, the range of correlations between the test scores and the simulated teacher judgment variable is mathematically constrained. Essentially, since we are dealing with a multiple correlation of pretest and teacher judgment with posttest and the pretest-posttest correlations are "given" in the empirical data, limits are placed on our ability to realize desired correlations between teacher judgments and pretest and posttest scores respectively.[2] The limits that were achieved (.2 and .6) are certainly wide enough to meet the needs of this study and representative of likely relations in the field.

It is very important to understand that the simulated correlations are used only to model the reallocation process in general. No claim is made (nor is one necessary to the argument) that the real-world correlations in any given Title I location between teacher judgments about children's needs for service and achievement test scores as indicators of those same needs follow anything like the structure we have built into our simulation. What is claimed is that the structure as simulated produces candidate cases for reallocation that cover the gamut of such cases that could (and do) occur in various locations in the nation.

To simplify somewhat, any student initially selected into the treatment on the basis of a pretest score should be considered as a potential candidate for reallocation on the basis of teacher judgments or local administrative policy. Similarly, all students initially excluded from the treatment group on the basis of the pretest should be considered as candidates for being moved into the treatment group. Figure 2 demonstrates this by exhibiting a scatterplot of the simulated teacher judgments on pretest scores for the students transferred out of the treatment group in one of the analyses to be presented subsequently. In this case, where the correlation of the teacher judgment

---

[2]We are indebted to A.O.H. Roberts for working out the mathematical solution to the problem of generating a third column of data having specified correlations with each of two existing columns of data, and to B.L. Stewart for performing the necessary computer programming.

with the pretest is much lower than its correlation with the posttest, the full range of pretest scores (below the cutoff) is represented in the group of students. Twenty-five students (slightly more than one-third of the total) had pretest scores between 31 and 36 NCEs, with five of those scores falling immediately below the cutoff (36.5 NCEs). Thirteen of the 72 scores are at or below the tenth NCE, indicating that students with extremely low pretest scores can be, under the conditions specified, transferred out of the selected group. Figure 3 contains a similar scatterplot for those students moved out of the original non-selected group and into the treatment group. These students, whose teacher judgment scores were quite low (indicative of substantial need for Title I services), also exhibit pretest scores covering the range from the cutoff to the upper boundary.

The results shown in Figures 2 and 3 illustrate that our simulation was successful in identifying students throughout the full range of pretest scores (above and below the cutoff) as candidates for reallocation. While some may argue that students whose pretest scores are quite proximal to the cutoff score are the "real" candidates for transfer, others may argue that teachers are less likely to disagree with scores close to the cutoff and, hence, students having the relatively more extreme scores would, in fact, be the "real" candidates for movement in and out of the treatment group. The simulation used in this study, of course, accommodates both of these specific cases and, further, accommodates the general case wherein teachers may disagree with any child's test score and act upon that disagreement.

Gain estimates. The norm-referenced model incorporates the equipercentile assumption in its estimation of program impact (gain). The equipercentile assumption holds that students will tend to maintain their percentile standing from pretest to posttest in the absence of special educational treatments such as Title I. The equipercentile assumption has been examined and found to be reasonable in earlier research (Wood, 1980). Where two separate measures are used to select

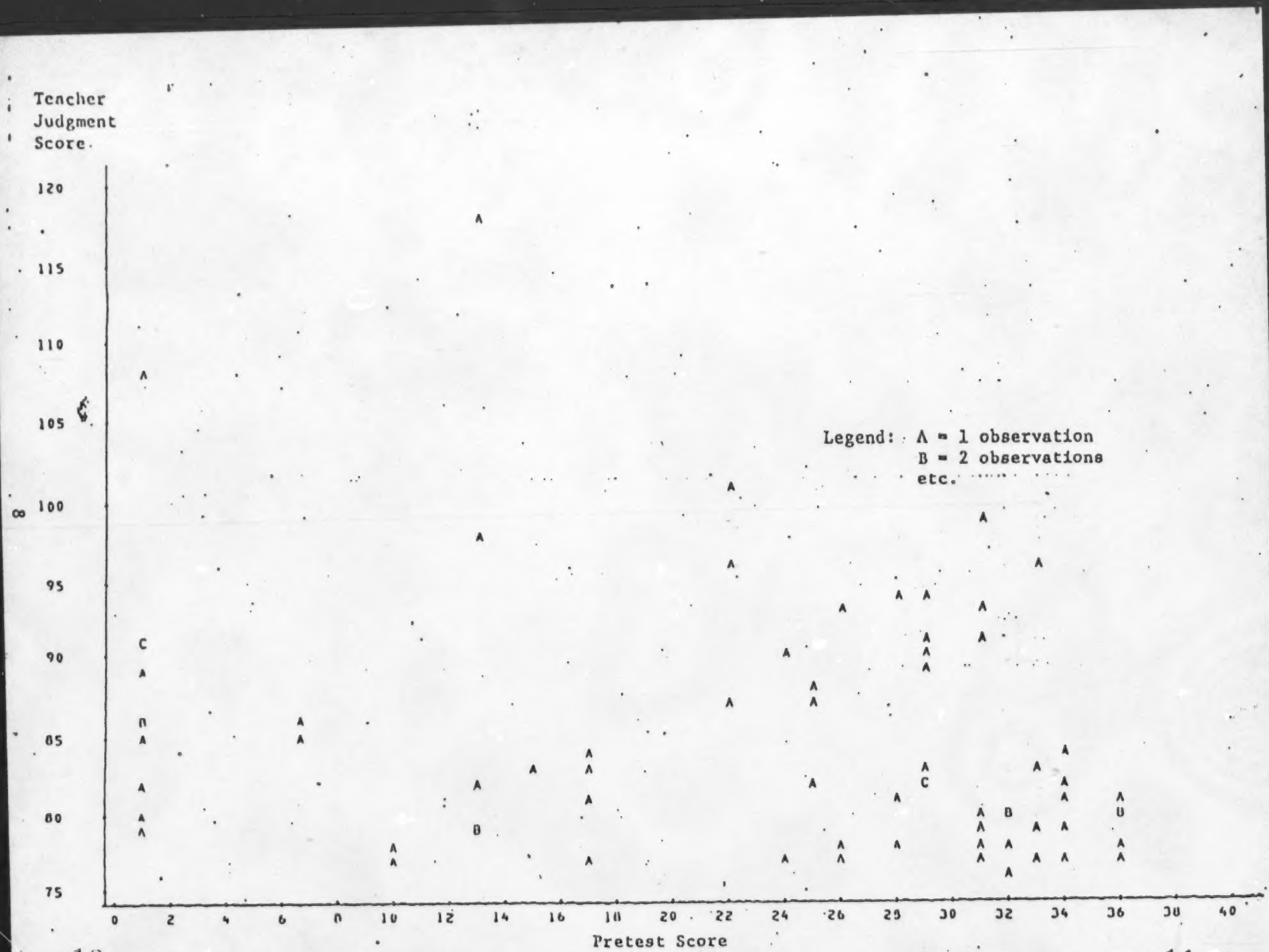

Figure 2. Plot of teacher judgment score against pretest score for students moved into the nonselected group.

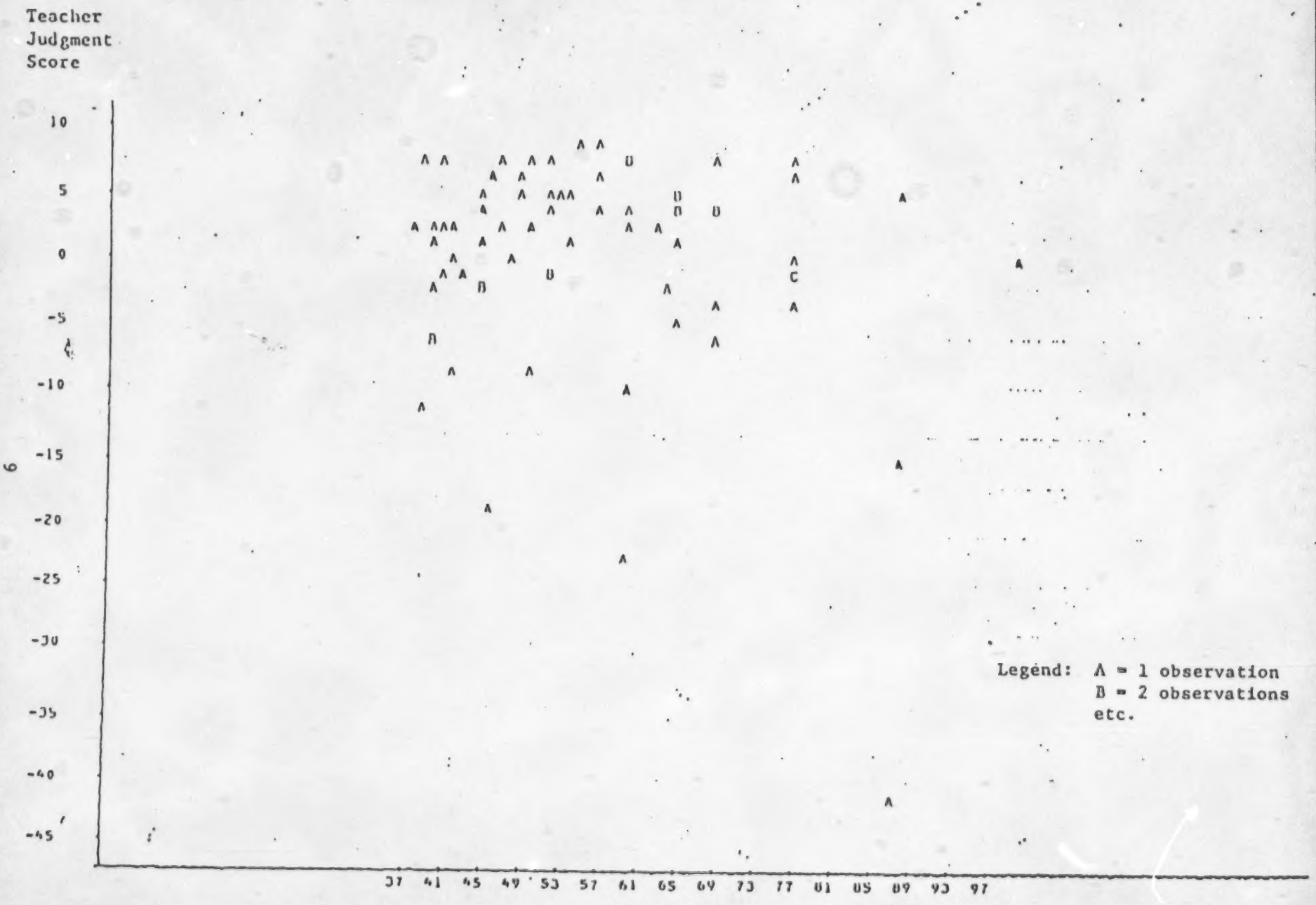

Figure 3. Plot of teacher judgment score against pretest score for students moved into the selected group.

and pretest students, the gain estimates, based on the equipercentile assumption, are simply posttest NCE scores minus pretest NCE scores. Of course, this is not an actual "gain" score; rather it is the difference between the observed score and the no-treatment expectation or baseline generated from the equipercentile assumption. Where the pretest score is used both to select project participants and to mark their initial status for use in estimating impact (gain), a correction to the pretest score is required. It is then the corrected pretest NCE that is subtracted from the observed posttest NCE for the purpose of assessing the gain.

In order to examine the effects of reallocating students between the selected and nonselected groups without the administration of a second pretest we generate three separate gain estimates under each of the three conditions of reallocation. The three estimates are as follows:

- Corrected for regression-effect error without reallocations (original selected group),
- "Corrected" after reallocations,
- "Corrected" after reallocation without adding replacement students to selected group.

The corrections referred to in the titles of the gain estimates are those for measurement errors in the pretest (Roberts, 1980). The actual correction formula was shown earlier. The estimates that involve reallocating students after the pretest cover both cases where new students are moved into the selected group to replace those who are moved out and where new students are not moved into the selected (the without replacement correlation).

The benchmark is the first gain score reported. This gain results from a legitimate application of the norm-referenced model with strict selection on the pretest and a correction to the pretest mean for the regression-effect error. The effect of the reallocation process under its three conditions can then be compared to the results under the benchmark condition. The reliability-based correction uses the pretest-posttest correlation of the total group of noncompensatory education students at grade 2 as the estimate of the appropriate test-retest reliability. Because the data for the entire group are available at both pretest and posttest, these simulations circumvent the difficulties in obtaining the appropriate reliability estimate for use in equation (1). Those difficulties are discussed at length by Roberts (1980). We should note that we could have opted to use the between-level correlations that are available as a part of the SES year-one data base as an estimate of the test-retest reliability. Since the between-level correlations were only trivially higher than the pre-post correlations (.800 vs. .785), we believe that choosing the pre-post correlation as the reliability estimate best serves the purposes of the simulations by allowing for very slightly more unreliability in the pretest means, thereby making the simulation somewhat more realistic vis-a-vis a system such as TIERS.

Aggregation levels. Separate analyses are presented for two levels of data aggregation. First, the effects of the reallocations are investigated in large-scale simulations, which are designed to estimate the effects that would occur in the nationally aggregated data. Second, simulations of the effects in small projects are reported, in order to illustrate the error that may be expected in local projects.

Results

Large-Scale Simulations

The results of the large-scale simulations are described separately for the two reallocation proportions studied. The three conditions under which the reallocations take place (i.e., the pattern of correlations of teacher judgment with pre- and posttest scores) are treated serially within the two presentations.

Effects with five percent of the students reallocated. The results of these analyses are summarized in Table 1. Table 1 displays the three gain estimates as described above that were obtained under each of the reallocation conditions (as determined by the correlations). The bias (i.e., deviations from the benchmark estimate) are shown in parentheses in the table. The resulting gain estimates under Condition A, in which the teacher judgment variable correlates substantially less with the pretest (.2) than it does with the posttest (.6), can be seen in the first column of Table 1. The summary test-score data, from which the gain estimates are generated, are shown in column "A" of Table 2. Note that the gain estimates for the original selected group do not change as a function of the reallocation condition: therefore, the same benchmark for comparison can be used under all three conditions.

Although not shown in the table, the gain estimate for the original selected group without employing the pretest score correction is a biased value of +4.91 NCEs. Once the pretest score (and thus the no-treatment expectation) is corrected using equation (1), the apparent gain is reduced to the negative value of -1.06 NCEs as shown in the table. The difference between these two estimates of nearly 6 NCEs is due, of course, to the regression-effect error associated with using a single measure for both selection and pretest purposes, and corresponds well with earlier estimates of the amount of bias in national aggregate results caused by selection on the pretest (Klibanoff, 1980).

Table 1

Three Estimates of Achievement Gain (in NCEs) for Three Levels of Correlation Between Teacher Judgment (Reallocation) and Pretest and Posttest Scores With a 5% Reallocation Proportion

Gain Estimate	Condition[a] (Correlation of Teacher Judgment with Pre- and Posttest) A[b]	B[c]	C[d]
1 Corrected for regression-effect error without reallocations (original selected group)	-1.06 (-)	-1.06 (-)	-1.06 (-)
2 "Corrected" after reallocations	-3.42 (-2.36)	-1.67 (- .61)	+ .49 (+1.55)
3 "Corrected" after reallocation without adding replacement students to selected group	-2.17 (-1.11)	-1.33 (- .27)	- .35 (+ .71)

[a]Estimates of bias (i.e., deviations from the benchmark) are shown in parentheses.

[b]Correlation of teacher judgment with pre- and posttest = .21, .61.

[c]Correlation of teacher judgment with pre- and posttest = .60, .60.

[d]Correlation of teacher judgment with pre- and posttest = .60, .20.

Table 2

Summary Test Score Data in NCEs for the Total Analysis Group and Six Subgroups Associated with the Reallocation of Students Following the Pretest Under Three Conditions, 5% Reallocation Proportions

	Nature of the Reallocation of Students							
	A[a]			B[b]			C[c]	
Subgroup	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
					Pretest				
Total Group	6449	51.74	20.26	6449	51.74	20.26	6449	51.74	20.26
Initially Selected	1443	23.97	10.15	1443	23.97	10.15	1443	23.97	10.15
Initially Non-Selected	5006	59.74	14.59	5006	59.74	14.59	5006	59.74	14.59
Transferred out of Selected	72	23.35	11.41	72	28.67	7.87	72	30.81	5.60
Transferred into Selected	72	54.93	14.78	72	46.53	7.99	72	46.47	8.44
New Selected	1443	25.55	12.36	1443	24.86	11.25	1443	24.25	11.28
New Non-Selected	5006	59.29	15.17	5006	59.48	14.98	5006	59.52	14.90
					Posttest				
Total Group	6449	51.02	19.94	6419	51.02	19.94	6449	51.02	19.94
Initially Selected	1443	28.88	13.48	1443	28.88	13.48	1443	28.88	13.48
Initially Non-Selected	5006	57.40	16.67	5006	57.40	16.67	5006	57.40	16.67
Transferred out of Selected	72	49.69	11.96	72	37.88	11.86	72	20.70	11.43
Transferred into Selected	72	27.25	14.44	72	39.60	13.71	72	64.03	16.59
New Selected	1443	27.76	12.74	1443	28.97	13.63	1443	31.04	15.57
New Non-Selected	5006	57.72	16.27	5006	57.38	16.68	5006	56.78	17.16

[a]Correlation of Teacher Judgment with Pre and Posttest = .21, .61

[b]Correlation of Teacher Judgment with Pre and Posttest = .60, .60

[c]Correlation of Teacher Judgment with Pre and Posttest = .60, .20

Examining the effects of the student reallocation, we can see that, under Condition A, underestimation of the gain (as compared to the benchmark) does result. When equation (1) is applied to the pretest mean of the new selected group, thereby accomplishing a partial correction (adjustment) for errors of measurement in the pretest, the negative gain is increased in size from -1.06 to -3.42 NCEs, a difference of more than 2 NCEs.

The third gain estimate shown in Table 1 is analogous to estimate 2 for the possible case where the new selected students (after the reallocation) are excluded from the evaluation sample. That is, for evaluation purposes, the reallocations are made without replacement. In this variant, the partially corrected (adjusted) gain is estimated to be -2.17 NCEs. This partially corrected estimate is a larger negative value than obtained when the correction formula is used. The practice of not replacing students who were moved out of the selected group reduces the bias associated with the reallocation, but does not eliminate it completely.

The second column of Table 1 supplies the gain estimates obtained under Condition B which establishes equal correlations for the teacher judgment variable with both pretest and posttest. Although the gain estimates for the groups after the student reallocation are closer to those before the reallocation than in Condition A, the results still appear to indicate some bias. The partially corrected gain shows a loss of .6 NCEs more than the corrected value. For the case of reallocation without replacement, the gain is closer to the original estimate than is the case that includes replacement students. In fact, it differs only slightly (.27 NCEs) from the value exhibited for the original selected group. Nevertheless, some negative bias appears to be the case.

The third column of Table 1 presents the estimated gains under Condition C. In this condition, the correlations reverse the pattern of Condition A. The teacher judgment variable now correlates more highly with the pretest than with the posttest. These correlations are intended to reflect the effects of reallocating students when the reallocations are based on judgments that are not good measures of students' achievement status. Such judgments may be reflecting again the fact that failure to maintain the sharp cutoff for selection yields biased results. However, the bias under this condition is positive rather than negative, although it is of the same order of magnitude as observed in the results for Condition A. The corrected value for the original selected group of -1.06 NCEs is one-and-a-half NCEs lower than the partially corrected estimate of +.49 NCEs obtained following the reallocation of students. Where no replacements are made for students transferred out of the selected group, bias is somewhat smaller than in the case where replacement students are assigned to the selected group. In this simulation, for example, the partially corrected gain excluding reallocated students from the evaluation sample is -.35 NCEs compared to the obtained +.49 NCEs where replacements are made.

To summarize results for the set of simulations, it appears that any reallocation of students between the selected and nonselected groups will introduce bias into the evaluation results. The data shown in Table 2 starkly illustrate how the bias is introduced under each of the three reallocation conditions. Table 2 displays the summary test-score data for the total group of students and six subgroups that are associated with the reallocation and norm-referenced evaluation procedures. The subgroups for which data are shown are those initially drawn into the selected and nonselected groups solely on the basis of pretest scores and the two groups of reallocated students (those moved out of the original selected group and those moved into the original selected group). The final set of statistics consists of the same summary data for the new selected and nonselected groups.

In Condition A, the pretest mean of the 72 students transferred out of the selected group is 23.35. That mean is rather similar to the mean of the total selected group (23.97), as is the mean pretest score of the new treatment group, which includes replacement students. The posttest mean for the 72 students transferred out of the selected group is 49.69, indicating a rather dramatic regression toward the mean of the group from which the initial selection was made. Clearly, negative errors of measurement were preponderant in the scores of students moved out of the original selected group. This rather substantial gain, resulting from the selection bias, is "lost" to the project once these students are transferred out, even when the original group's mean is corrected for unreliability. When replacements scoring above the original cutoff score are brought into the selected group, the effect is further accentuated. The pretest mean for these new members of the treatment group is 54.93, which has the effect, of course, of raising the treatment group's pretest mean by a small amount (1.58 NCEs). At posttest, however, this group of 72 new students has a mean score of 27.76 NCEs, indicating a dramatic regression effect. When the reallocations include replacement students, the twofold effect on project gain explains the fact that the error is larger than that found in the simulations that do not include the replacement students. In this simulation, the net difference between the two groups is -27.68 NCEs.

The results under reallocation Condition B, while still exhibiting negative bias, are a good deal less extreme than those obtained in the Condition A simulation. In this simulation, the pretest mean for the group transferred out of the selected category is 28.67 (5.32 NCEs higher than the corresponding group in Condition A). Its posttest mean is 37.88, a value that is indicative of the fact that on balance less extreme cases (students) have been subject to the reallocation process. Although the results reflect a lesser degree of bias, some bias, nevertheless, remains. The students transferred into the selected group show a gain of -6.93 NCEs, for a net difference of 16.14 NCEs. Even where a fairly small percentage of the original selected and nonselected groups

are reallocated, the reallocation results in a decided negative impact on the gain estimate.

In Condition C, where we posit a different underlying nature for the reallocation process--one that mimics either poor judgments by teachers or judgments that appear to measure some student characteristics other than achievement--similar degrees of bias are evident in the results. However, in this condition, the bias is in the positive rather than negative direction. The pretest mean of the students transferred out of the selected group is higher than those observed under the preceding two conditions and is quite substantially higher than that under Condition A (7.46 NCEs or more than one-third of a standard deviation). The resulting deflation of the new treatment group's pretest mean is enough, even without adding replacement students, to yield an inflated estimate of the project's impact, as can be seen in the results of Table 1. Where replacement students are added the effect is again larger. In the simulation, the replacement students have a lower pretest mean (46.47) compared to 54.93 (in Condition A) and have a posttest mean of 64.03, exhibiting substantial positive gain. Again, when replacements are made the inflation is accentuated.

Effects with ten percent of the students reallocated. The above analyses were repeated with a larger group of students subjected to reallocation under each of the three conditions. The results are shown in Tables 3 and 4. Table 3 exhibits the gain estimates and bias (similar to Table 1); Table 4 provides the summary data analogous to those of Table 2. An inspection of the data in Table 3 clearly shows that an increasing amount of bias is associated with increasing proportions of students subjected to reallocation.

Under Condition A, the gain estimates are depressed relative to those obtained for the selected group prior to the reallocations. When the partially corrected estimates (gains 2 and 3) are compared to the corrected estimate before the reallocation (gain 1), the post-reallocation estimate including replacement students (gain 2) is

Table 3

Three Estimates of Achievement Gain (in NCEs) for Three Levels of Correlation Between Teacher Judgment (Reallocation) and Pretest and Posttest Scores With a 10% Reallocation Proportion

Gain Estimate	Condition[a] (Correlation of Teacher Judgment with Pre- and Posttest) A[b]	B[c]	C[d]
1 Corrected for regression-effect error without reallocations (original selected group)	-1.06 (-)	-1.06 (-)	-1.06 (-)
2 "Corrected" after reallocations	-5.10 (-4.04)	-2.21 (-1.15)	+1.61 (+2.67)
3 "Corrected" after reallocation without adding replacement students to selected group	-3.16 (-2.10)	-1.52 (- .46)	+ .27 (+1.33)

[a]Estimates of bias (i.e., deviations from the benchmark) are shown in parentheses.

[b]Correlation of teacher judgment with pre- and posttest = .21, .61.

[c]Correlation of teacher judgment with pre- and posttest = .60, .60.

[d]Correlation of teacher judgment with pre- and posttest = .60, .20.

Table 4

Summary Test Score Data in NCEs for the Total Analysis Group and Six Subgroups Associated with the Reallocation of Students Following the Pretest Under Three Conditions, 10% Reallocation Factor

	Nature of the Reallocation of Students							
	A[a]			B[b]			C[c]	
Subgroup	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
					Pretest				
Total Group	6449	51.74	20.26	6449	51.74	20.26	6449	51.74	20.26
Initially Selected	1443	23.97	10.15	1443	23.97	10.15	1443	23.97	10.15
Initially Non-Selected	5006	59.74	14.59	5006	59.74	14.59	5006	59.74	14.59
Transferred out of Selected	144	20.70	12.14	144	29.72	6.14	144	30.01	5.87
Transferred into Selected	144	55.55	14.38	144	46.73	8.03	144	47.97	8.68
New Selected	1443	27.45	13.98	1443	25.67	12.30	1443	25.76	12.56
New Non-Selected	5006	58.74	15.92	5006	59.25	15.27	5006	59.23	15.26
					Posttest				
Total Group	6449	51.02	19.94	6449	51.02	19.94	6449	51.02	19.94
Initially Selected	1443	28.88	13.48	1443	28.88	13.48	1443	28.88	13.48
Initially Non-Selected	5006	57.40	16.67	5006	57.40	16.67	5006	57.40	16.67
Transferred out of Selected	144	45.31	11.68	144	37.57	14.12	144	21.69	11.02
Transferred into Selected	144	32.17	14.92	144	39.43	14.32	144	62.52	14.75
New Selected	1443	27.57	12.76	1443	29.07	13.63	1443	32.96	16.81
New Non-Selected	5006	57.78	16.16	5006	57.35	16.73	5006	56.23	17.60

[a]Correlation of Teacher Judgment with Pre and Posttest = .21, .61

[b]Correlation of Teacher Judgment with Pre and Posttest = .60, .60

[c]Correlation of Teacher Judgment with Pre and Posttest = .60, .20

slightly more than 4 NCEs larger than the corrected estimate for the original group. When the partial correction is effected without adding replacement students to the evaluation sample, the gain is still more than 2 NCEs more negative than the benchmark value.

As was the case in the simulation using a five percent reallocation, the results under Condition B, while still negatively biased, are less extreme than those for Condition A. The partially corrected estimates are 1.15 and .46 lower than the benchmark for reallocation with and without replacement respectively.

In Condition C, the results are again positive and yield larger gains than in the five percent simulation. The deviation from the benchmark is 2.67 NCEs in the case where replacement students are added and is 1.33 NCEs where replacement students are excluded from the evaluation sample. Both represent overestimates of treatment impact.

The results where ten percent of the students from each region of the original test score distributions are reallocated exhibit the same pattern as those for the five percent simulation, but with larger biases. The causes can be seen in the summary test-score data shown in Table 4. In both Conditions A and B, the students transferred out of the selected group exhibit substantial gains in their mean scores from pretest to posttest (24.61 NCEs in Condition A and 7.85 NCEs in Condition B). Although these gains are not vastly larger than those made by similar groups in the five percent simulation (26.34 and 9.21, respectively), the larger number of students involved, of course, has a concomitantly greater effect on the eventual group mean. The students who are transferred into the selected group show a fairly large loss from pretest to posttest, thereby enlarging the bias. In Condition C, the pattern is reversed with large gainers being moved into the treatment group and large decliners being moved out.

Effects at the Project Level

The preceding results were intended to shed light on the overall issue of the fuzzy cutoff by exploring the effects in a large-scale simulation. The large-scale simulation accomplishes two ends. First, it allows us to study the problem under conditions which minimize the effects of random error in the data that could cloud the results. Second, it provides us with good estimates of the effects that would be evident in the national aggregation aspect of TIERS. Large-scale simulations do not, however, give us very good estimates of the size of errors that should be expected in local projects of various sizes. The following analyses address the issue at the local or project level.

We used data from the five percent reallocation simulation in which 72 students were transferred out of the selected group and 72 students were transferred in. The two groups of 72 students then represent a pool from which reallocations can be made. The data were treated as if they had come from projects serving 10, 20, and 30 students. The pretest means of the projects were assumed to be 23.97 NCEs, the same mean obtained for the initially selected group in the large-scale simulation. Students with extreme positive or negative gain scores in the original group of 72 transferred out students were then removed from their "projects," both with and without replacements by students from the other group of 72.

The results of the analysis are shown in Table 5. The data were those from the five percent simulation performed under reallocation Condition A. Not surprisingly, the smaller the project, the greater the impact of moving even one student. In fact, these results indicate that even such a minimal change in the constitution of a 30-student project can have a negative impact of from 2 to over 4 NCEs depending upon whether or not replacement students are brought into the project. Since many Title I projects have even fewer than 30 students, the potential biases are dramatically large--on the order of 6 to 13 NCEs in a 10-

student project. Obviously, the effect for any single project depends on which students are moved.

Table 5

Gain Estimates Following Student Reallocation in Three Sizes of Title I Projects

Number of Students Transferred/ Project Size	Gain Estimate (Bias)		
	Original Group	With Replacement Students	Without Replacement Students
1/10	-1.06	-13.71 (-12.65)	-7.52 (-6.46)
1/20	-1.06	- 7.39 (- 6.33)	-4.12 (-3.06)
1/30	-1.06	- 5.28 (- 4.22)	-3.06 (-2.00)

The question of interest here is, "What would be the effects in those projects that moved the most extreme cases?" The results shown in Table 5 indicate the maximum values that could be obtained using the data for transferred students generated as a part of the earlier simulation. They do not, however, represent the maximum theoretical bias that could result from student reallocation. The "student" who was transferred out of the treatment group had a pretest of 1 NCE and a posttest score of 69 NCEs. The replacement student had a pretest score of 87 NCEs and a posttest score of 10 NCEs. These two "students" exhibited the largest changes from pretest to posttest in the empirical data, but it is conceivable that students with even larger gains (and declines) from pre- to posttest could be reallocated. Using these extremes in a 10-student project having the same pretest, posttest, and reliability values as those obtained in our simulation (pretest of 23.97, posttest of 28.88, reliability for the total group from which the selection was made of .785), even larger biases would result. In the example, the resulting bias could exceed 17 NCEs on the negative side. Bias on the positive side, because it is constrained by the original cutoff score, could reach nearly 15 NCEs in projects serving 10 students with one transfer moved out and replaced by another student. Even without adding an extreme replacement student, the bias could approach 12 NCEs on the negative size and 4 NCEs on the positive side.

Discussion

This research was directed at exploring the ramifications at both the aggregate and project levels of a failure to maintain a sharp cutoff for the selection of Title I participants when the pretest score is used both as a selection device and initial-status indicator. On the whole, the results suggest that reallocating students after the pretest will, even in conjunction with the correction for pretest unreliability, bias the evaluation results. At the local level, the biases are potentially quite large. In projects serving very few students, the potential bias from moving even one student approaches two-thirds of a standard deviation (i.e., 14 NCEs). Of course, if students showing less extreme changes from pretest to posttest were exchanged, the size of the bias would be concomitantly smaller, but the potential for large biases at the local level remains quite real. At the national level, the bias will, of course, be much smaller. But little comfort can be taken in that result, since the bias may be positive or negative depending upon the actual basis on which the reallocations are made.

Of course, the usual conclusion when both positive and negative estimation errors are found is that the estimation procedure is unbiased or has only a small bias depending upon the predominant deviation of the observed errors. In this case, however, that conclusion probably does not follow. On the one hand, our simulations do not inform us at all about the relative prevalence of reallocation procedures that lead to negative and positive biases respectively. That is, our correlational model says nothing about which of the three reallocation conditions best represents the picture at the level of the national aggregate. It would be convenient to assume that Condition B, which posits that the teacher judgment variable correlates equally with pretest and posttest scores, would provide the most viable description of the process. In that case we could presume a fairly small negative bias to the procedure that could be "taken account of" at the aggregate level. However, there is little a priori reason for making that assumption, since that would put

us in the position of having to assume that the reallocations have no validity at all. Given that local personnel, especially teachers, can be quite well informed about their students' achievement levels, the assumption of zero validity is probably untenable.

The intent of this study was to examine the effect of the so-called fuzzy cutoff in conjunction with the correction for regression effect error when the same measure is used for both selection and pretest purposes. The results indicate that, depending upon the nature of the process by which the reallocations are made, the correction formula is probably not an effective correction. In all cases, applying the correction formula (equation 1) following the reallocation of students reduced the absolute size of the bias but did not effect as adequate a correction as that accomplished when used after a strict cutoff. The failure of the correction formula to compensate fully for the bias caused by the reallocation is not surprising, even with the large population and sample sizes used.

The sharp cutoff serves the function of ensuring that selection into the treatment group is based on only one variable, pre-treatment achievement level. Because that level is, of course, assessed by a measure that is not perfectly reliable, regression toward the mean occurs. The correction, using the test-retest reliability for the total group, is applied and serves the purpose of a "second" pretest. The retesting restores the balance between errors of measurement and true scores, thereby allowing an accurate measurement of the pretest status of the treatment group from which the no-treatment expectation can be generated. (See also Nesselroade, Stigler, & Baltes, 1980.) The balance that is restored either by retesting or applying the correction is not achieved in the case where reallocations are made. Applying the correction in such cases reduces the magnitude of the bias but does not eliminate it.

In the case of TIERS, where the purpose is always to select a low-achieving group for treatment, the correction will adjust the pretest score in the correct direction but will invariably lead to underestimates of project impact where the reallocations have real validity (i.e., Condition A in the simulations) and will invariably yield overestimates of project impact where the reallocations are not valid. Unfortunately, the different processes that underlie the reallocations do not allow us to draw the conclusion that the inaccurate estimates will balance out in the aggregation of the results at the state and national levels.

The fuzzy cutoff studied in this research differs substantially from the more common conception popularized by Campbell (1969). In that more usual formulation the fuzzy area of the bivariate scatter-plot of posttest on pretest is constrained to a region close to the original cutoff score. Figure 3 shows this graphically. The horizontal axis of Figure 3 represents the pretest scores. The shaded areas of the figure represent students who would be candidates for transfer across the cutoff boundary. Contrast Figure 3 with the view given in Figures 1 and 2, wherein students throughout the pretest range were considered as potential candidates for reallocation. Even if the candidates for reallocation were restricted to those students whose pretest scores fall within the shaded area of Figure 3, a considerable degree of bias could still be introduced into a local evaluation. Consider our simulated 10-student project again. By moving one student from each side of the cutoff score (using actual data from our earlier simulation), gain estimates of -7.34 and -5.36 were obtained for cases with and without replacement, respectively. Both of the obtained estimates are substantially larger than the benchmark estimate of -1.06 NCEs.

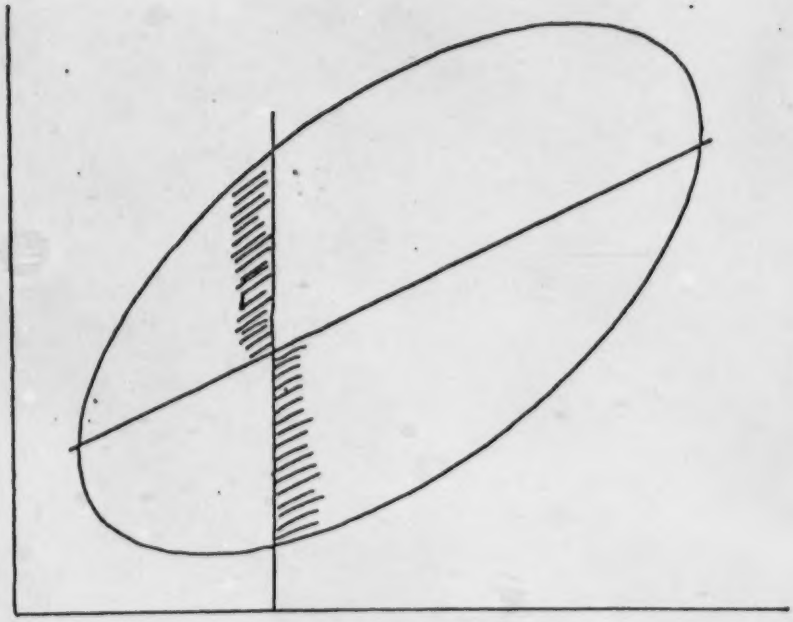

Figure 4. The classical fuzzy cutoff.

The fact that we are still able to generate rather large biases by exchanging cases that fall immediately adjacent to the cutoff is probably related to the nature of the correlation of true scores and error within each of the two subgroups created by the cutoff. The negative correlation between the two test-score components tends to be greater nearer the cutoff score. The greater negativity of the true-score-to-error correlation near the cutoff suggests, as shown in the simulation, that large biases can be expected to result from reallocations even if the reallocations are restricted to some pretest band proximal to the original cutoff.

Because of the central importance of the norm-referenced model, technical studies directed at its implementation, operation, and outcomes are an important part of the evolutionary development of TIERS. The norm-referenced model is by far the most widely used of the three sets of TIERS models. Although both from a technical viewpoint and from an implementation viewpoint the norm-referenced model is less complex than either the comparison group or regression models, any simplification of its procedures and any reduction of burden (either cost or student time) would enhance the value of the model and of the system. The provision to allow a single test to be used for both selection and pretest purposes is one major simplification recently made to the model.

Allowing for reallocations of students following a single selection/pretest would represent another refinement to the norm-referenced model that would enhance its implementation at the local level. However, the results of this research strongly contraindicate such an allowance. The errors at the aggregate levels ranged from .27 to 4.04 NCEs, depending upon the proportion of students moved, and whether or not new students were added to the project. Were the system to allow such reallocations, the most likely procedure would be to exclude new students from the evaluation sample and to apply the reliability correction. Under that scenario the observed errors ranged from .27 NCEs to 1.11 NCEs (absolute values) where five percent of the students were reallocated; under the ten percent reallocation the absolute magnitude of the errors ranged from .46 to 2.1 NCEs. Although not very large, such errors would represent a quite undesirable bias in the aggregated results. However, it is the rather large amount of bias that would occur at the local level that truly militates against allowing reallocation procedures. As we have seen, removing even one student from a small project brings an unacceptably large bias with it.

On the whole, results suggest that reallocating students after the pretest will bias evaluation results, even when reallocation is done in conjunction with a correction for pretest unreliability. For a small

project, the potential effect of moving even one student approaches two-thirds of a standard deviation (i.e., 14 NCEs). The amount of bias always depends on the relationships between those data used to make the reallocation decision (here referred to as teacher judgment) and the pretest and posttest scores of the reallocated student. This relationship is quite unpredictable and cannot be ascertained before the fact. Since there is little basis for assuming that the reallocation variable would correlate equally with pretest and posttest scores (the case resulting in minimal bias), reallocation after the pretest is not recommended for a norm-referenced evaluation.

It would appear that the two existing options will need to suffice. Local educational agencies who desire to use a single test to select and to pretest their students should strictly observe the initial cutoff score they employ in order to avail themselves of the correction formula. Those districts that wish to include the judgments of teachers or other personnel in the selection process will need to administer a separate, norm-referenced test to establish the pretest status of the selected group.

References

Campbell, D. T. Reforms as experiments. American Psychologist, 1969, 24, 409-429.

Klibanoff, L. S. Estimating the bias in reported Title I achievement gains for fiscal year 1978. Mountain View, CA: RMC Research Corporation, 1980.

Nesselroade, J. R., Stigler, S. M., & Baltes, P. M. Regression toward the mean and the study of change. Psychological Bulletin, 1980, 88, 622-637.

Roberts, A. O. H. Regression toward the mean and the regression-effect bias. New Directions for Testing and Measurement, 1980, 8, 59-82.

Wang, M. M., Hoepfner, R., Zagorski, H., Hemenway, J. A., Brown, D. S., & Bear, M. B. The nature and recipients of compensatory education. Report No. 5 from the Sustaining Effects Study. Santa Monica, CA: System Development Corporation, 1978.

Wood, C. T. The adequacy of the equipercentile assumption. Mountain View, CA: RMC Research Corporation, 1980.